Knowledge-based methodology in pattern recognition and understanding

Jean-Paul Haton

UNITÉ DE RECHERCHE
INRIA-LORRAINE

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
B.P.105
78153 Le Chesnay Cedex
France

Tél.:(1) 39 63 55 11

Rapports de Recherche

N° 691

KNOWLEDGE-BASED METHODOLOGY IN PATTERN RECOGNITION AND UNDERSTANDING

Jean-Paul HATON

JUILLET 1987

KNOWLEDGE-BASED METHODOLOGY IN PATTERN RECOGNITION AND UNDERSTANDING

Jean-Paul HATON

CRIN / INRIA-Nancy
B.P. 239
54506 Vandoeuvre les Nancy cedex, France

The interpretation and understanding of complex patterns (e.g. speech, images or other kinds of mono- or multi-dimensional signals) is related both to pattern recognition and artificial intelligence since it necessitates numerical processing as well as symbolic knowledge-based reasoning techniques.
This paper presents a state-of-the-art in the field, including basic concepts and practical applications.

1. INTRODUCTION

Artificial Intelligence (AI) is concerned with symbolic information processing such as encountered in human activities of reasoning, interpretation, decision making, etc. The field of AI is presently experiencing a rapid growth, both in research and development, especially through the techniques of knowledge-based systems. On the other hand, pattern recognition (PR) covers a set of numerical processing methods for the purpose of perception and pattern classification.

The interpretation of complex patterns - e.g. speech, images, acoustic or bio-medical signals - makes it necessary to combine AI and PR methods since it combines numerical processing techniques together with knowledge-based symbolic reasoning. The interpretation of patterns or signals can be roughly described as the progressive transformation of physical data coming from sensors into a symbolic and finally semantic description specific to each application.

The convergence of these two complementary domains of AI and PR is a characteristic of today's activities, necessitated by the design of increasingly sophisticated systems in advanced domains such as robotics, speech understanding, computer vision or image interpretation.

We present in this paper some problems and solutions associated with these activities. In section 2 the fundamentals of PR are recalled, with the two main approaches of statistical PR and structural PR. Section 3 presents the problem of pattern understanding as a knowledge intensive process. Main present issues are introduced in section 4 : production rule systems, object-oriented knowledge representation and reasoning. Section 5 is concerned with the important problem of incorporating various knowledge sources in the interpretation process. Finally section 6 presents the control structures and strategies that can be implemented in understanding systems. The basic concepts will be illustrated throughout the paper by practical examples from different fields of application.

LES SYSTEMES A BASES DE CONNAISSANCES EN RECONNAISSANCE ET COMPREHENSION DE FORMES

L'interprétation et la compréhension de formes complexes (par exemple parole, images ou d'autres signaux mono- ou multi-dimensionnels) sont liées à la fois à la reconnaissance des formes et à l'intelligence artificielle. Ces problèmes nécessitent en effet de faire appel à des techniques numériques de traitement mais aussi à des raisonnements symboliques fondés sur des connaissances.
Cet article présente un état de l'art du domaine, tant sur le plan des concepts de base que sur celui des applications pratiques.

2. BASIC PRINCIPLES OF PATTERN RECOGNITION

PR aims at simulating the perception capabilities of human beings such as vision or hearing. A similarity with AI lies in the fact that both are concerned with the study of human behaviour. A specific feature of PR is the fact that PR has to deal with data which come directly from the physical world through sensors (microphones, cameras, scanners, etc.). The basic problem consists in representing these physical signals in some adequate parametric space and then in transforming them into a symbolic form, for instance :

- from a matrix of pixels (representing an image) extract primitive patterns (vertex, segment, etc.) and identify an object,
- from a digitized speech wave compute a parametric representation (e.g. by spectral analysis) and identify a word.

Such activities necessitate the use of a dictionary of reference patterns that are stored during a preliminary learning phase. Figure 1 summarizes this process.

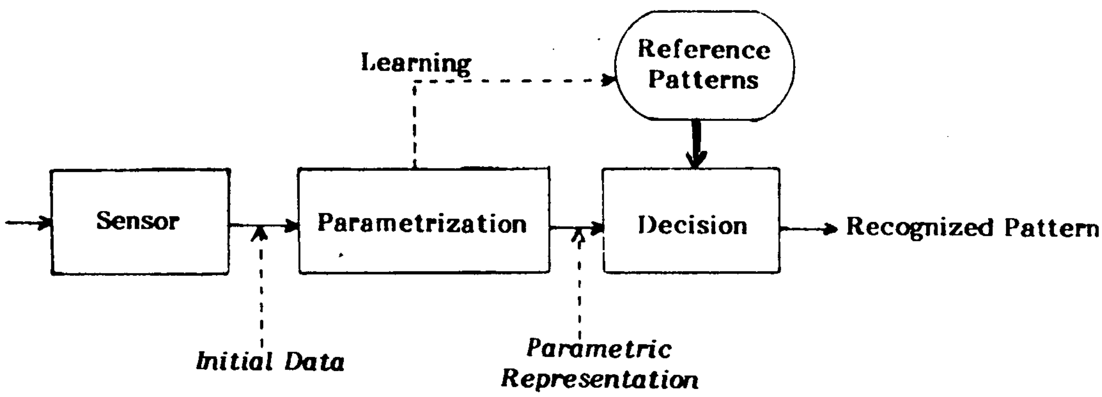

FIGURE 1
Principle of a pattern recognition system.

Three consequences are the result of manipulating data coming directly from the external world and these highly affect the overall organization of the system :

- undeterminism due to the fuzziness of the problem, to noisy and incomplete data, etc.,
- variability specific to the underlying physical phenomena,
- large amount of data to be processed (typically 100 000 bits for 1 second of speech, 4 000 000 bits for a TV image or a A4 typed page). That makes it necessary both to reduce these data by signal processing and/or data analysis techniques and to limit the complexity of algorithms.

PR methods try to reduce the undeterminism and variability of data according to two main, non exclusive approaches |1| |2| :

- **statistical methods** in which reference patterns are made up of vectors of mean parameters. Related decision rules consist of computing the neighbourhood of a pattern (e.g. the k NN rule) or of studying the position of a point from separating hypersurfaces in a multi-dimensional feature space. Stochastic models have also been widely used in coping with variability in domains like speech recognition (for instance Hidden Markov Models |3|). Statistical methods are efficient and general. However they are intrinsically limited by the fact that they rely on a mathematical modelling of the phenomena and therefore do not take into account the structure and the physical reality associated with patterns. A statistical approach is therefore not sufficient for complex tasks (e.g. understanding of spoken sentences, vision of a mobile robot) whereas they represent efficient solutions to simpler problems such as word recognition or simple object identification ;

- **structural or syntactic methods** are related to the structure of the patterns to be recognized and to their description as the assembly of primitive patterns. The recognition consists of the comparison of two structures that can be carried out by syntactic parsing techniques |4|, or by relaxation labelling |5|. Since these methods are basically deterministic and few tolerant to segmentation errors and noises some statistical capabilities have sometimes been added : stochastic grammars |6|, probabilistic relaxation. Presently available models of structural PR do not allow for the representation of the various knowledge sources that participate in the interpretation of patterns even though they offer better capabilities compared to statistical models.

A PR system is a necessary front-end to any AI system that must understand and interpret perceptual data. The overall performances in understanding are directly dependent upon the quality of this front-end (e.g. in speech understanding and computer vision). The field of PR is presently evolving towards greater complexity in the general framework of pattern understanding. This evolution can be found in speech recognition (from isolated and connected word recognition to sentence and dialog understanding) as well as in image processing (from object identification to scene analysis) or in optical reading (from character recognition to text interpretation). Purely procedural techniques become insufficient in solving such complex problems : it becomes necessary to take into account all available types of information and knowledge, especially with the AI techniques of knowledge-based systems.

3. PATTERN UNDERSTANDING : A PROBLEM OF PR AND AI

It has already been stated that the understanding of complex patterns or signals consists of extracting from rough initial data a semantic description that is used to carry out some predetermined task : control of a mobile robot, medical diagnosis based on radiological images, answer to an oral enquiry, etc.

This high-level description can only be obtained by a progressive transformation of the initial data through successive symbolic steps based on the use of a large variety of knowledge sources. At each step of the transformation there exist specific mechanisms which structure the information while keeping it under a minimum volume. That can be referred to as the simplicity principle in cognition, which states that patterns are analyzed by the human brain and stored under the simplest possible form. A PR system constitutes the first step in the process. The imbrication between PR and AI techniques is therefore very high.

In order to ensure the efficiency of a system it is necessary to formalize all relevant knowledge sources and to integrate them into an adequate architecture. Figure 2 illustrates the general principle of an understanding system in which several knowledge sources (KS), each with a specific activation mechanism (M) are cooperating together.

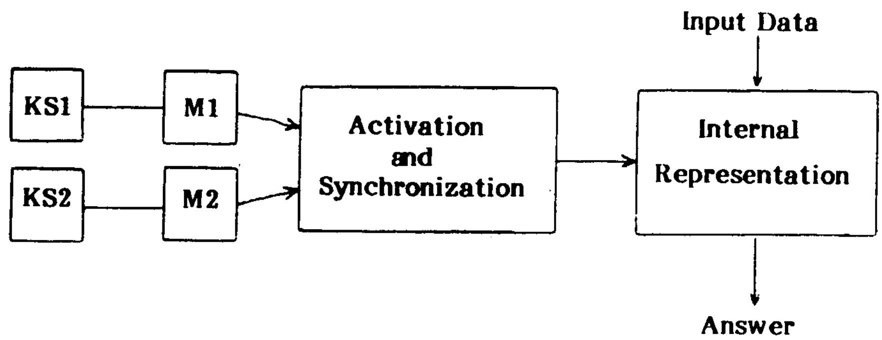

FIGURE 2
General architecture of an understanding system.

It is not only necessary to use as good and as complete knowledge sources as possible but also to use it adequately and efficiently. Section 5 will present some architectures that have been developed for this purpose. Figures 3 and 4 give an idea about the nature of knowledge sources which intervene in two typical domains of understanding : respectively, computer vision and speech understanding [7].

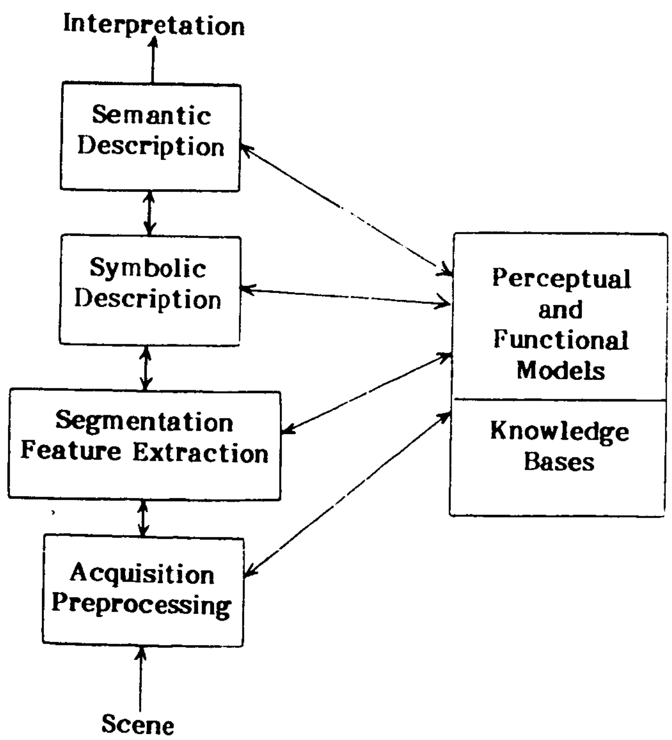

FIGURE 3
Principle of a computer vision system

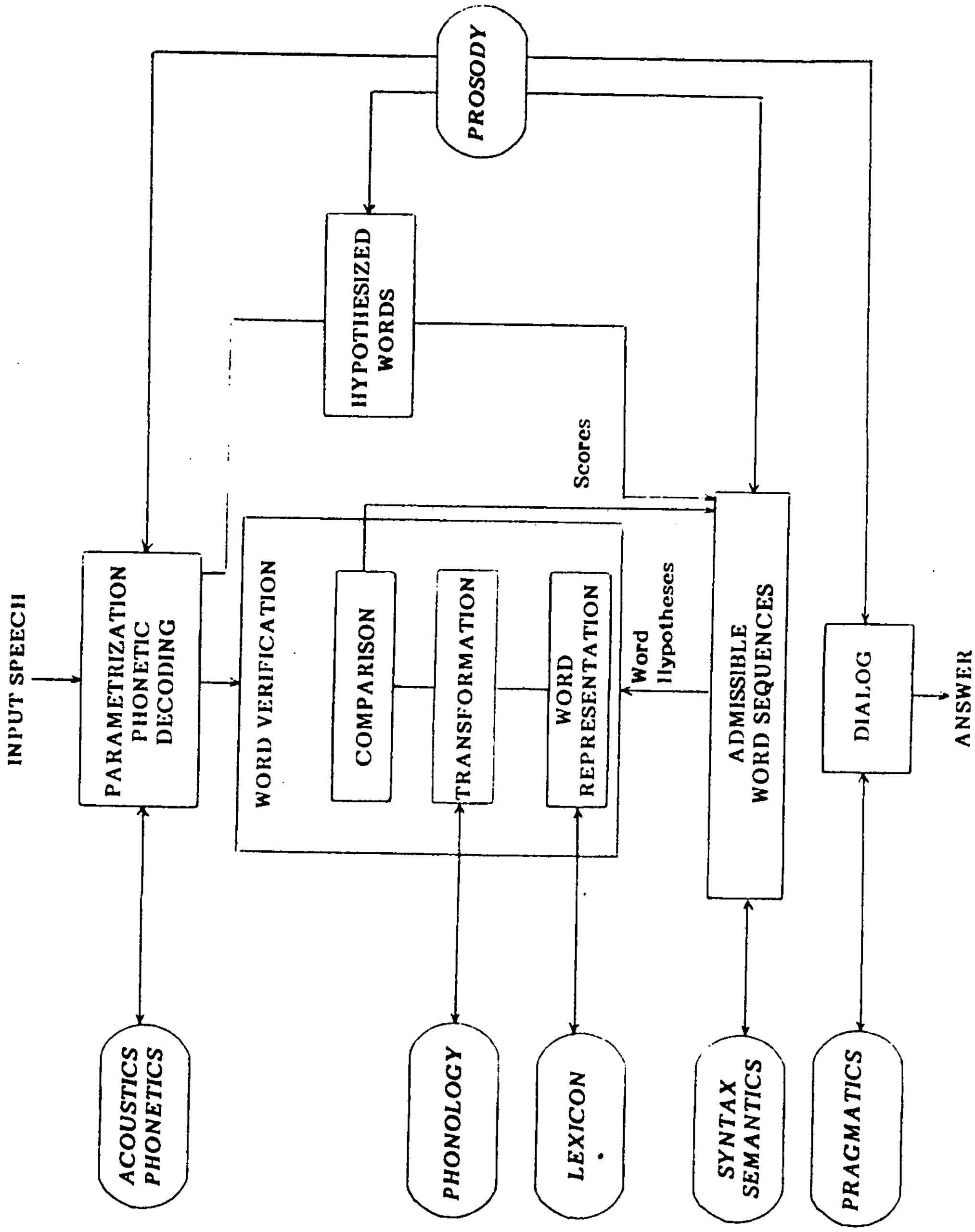

FIGURE 4
Processing levels and knowledge sources in speech understanding.

Figure 4 shows the central role of the lexical, or word level in speech understanding together with the importance of the acoustic-phonetic decoding of the speech wave that constitutes the front-end between the physical speech signal and the linguistic symbolic processors.

A similar situation can be found in computer vision where the interaction between low-level and high-level processing takes place at the level of volumes and objects which appear functionally similar to the words in speech processing.

The corresponding human process is basically carried out by hypothesis emission and controlled by the knowledge accumulated by learning. The paradigm of hypothesis-and-test, or prediction-verification presents some analogy with this mechanism. Figure 5 shows the principle of this reasoning scheme together with its practical materialization in the case of speech and vision. The basic idea is to emit hypotheses (e.g. the presence of a word or an object in the input data) according to available knowledge and according to the context and then to test these hypotheses on the current representation of data.

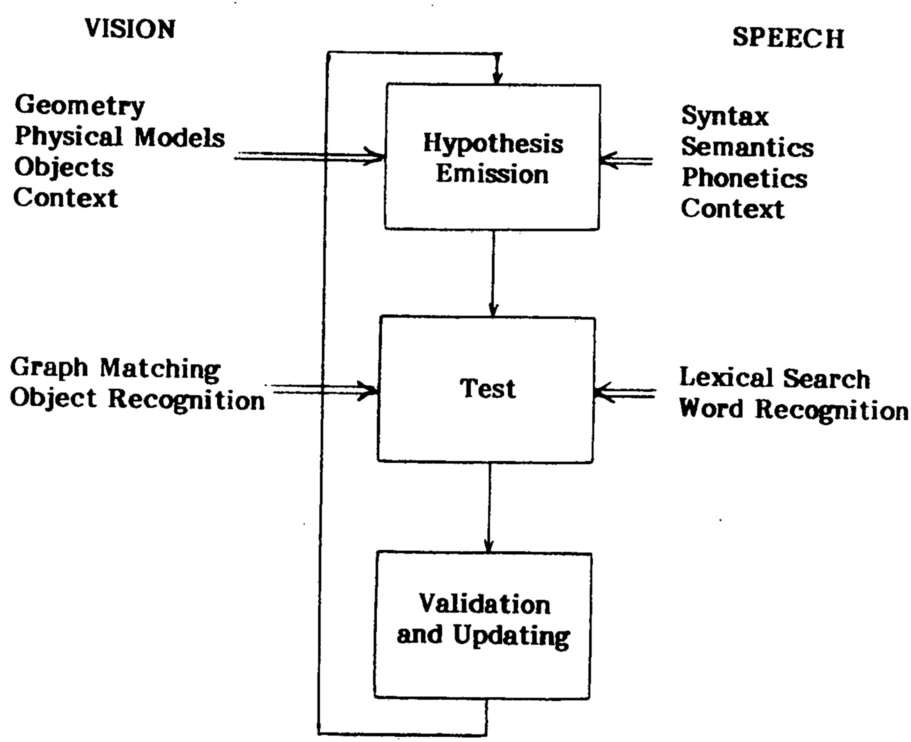

FIGURE 5
The Hypothesis-and-Test Reasoning Scheme.

As a matter of fact the analogy with the human brain is only superficial since human beings present a great ability of simultaneously processing a large number of hypotheses in such a way that these hypotheses influence each other and the final solution is consistent with input data and available knowledge and information.

Amongst the various kinds of knowledge that can be found in understanding systems, models play an important role. In speech recognition the lexicon stores word models with different pronunciations, sometimes precompiled. In scene and image analysis, objects are also described by models, especially under the form of graphs |8| and of semantic nets |9| |10|. Such models can be complemented by other formalisms like production rules (cf. section 4). Some generic models are used in vision in order to generate various classes of objects, for instance the 3D model of generalized cones used for recognizing objects on 2D images in the ACRONYM system |11|.

4. CURRENT ISSUES AND DEVELOPMENT

4.1. Rule-based Reasoning

The knowledge base of present expert systems is often partially or totally made up of production rules. Such rules represent "chunks" of knowledge in a Condition/Action formalism :

IF Condition THEN Conclusion/Action.

They can be used either in a bottom-up, data-driven way ("forward chaining") or in a top-down, goal-driven way ("backward chaining"). Production rules provide a rather natural means for representing the surface knowledge and heuristics that are used by a human expert, especially for diagnostic and data interpretation tasks. They have been used for various applications of pattern interpretation, sometimes in association with other representation schemes for deep knowledge :

Acoustic-Phonetic Decoding of Speech -
Acoustic-phonetic decoding is an operation of primary importance in automatic speech recognition. It consists in transforming the continuous speech signal into a sequence of discrete linguistic units (phones, phonemes, diphones, syllables). This is a very difficult operation, especially in a multi-speaker framework, and the results obtained are still limited. Usual methods are based on classical PR approaches with phonemic models (prototype patterns of speech spectra, vectors obtained by vector quantization of the speech wave, stochastic models, etc.). Such algorithmic "blind" methods can be advantageously complemented by some amount of rule-based reasoning, due to the contextdependent aspect of phonetic alterations. A method for formalizing these knowledge rules is to refer to the activity of speech spectrogram reading by expert phoneticians. Several studies have been carried out for English |12| and for French |13| |14| |15|.

The development of an expert system for phonetic decoding makes it necessary to solve specific problems generated by the characteristics of the speech signal and by the inherent undeterminism of the domain :
- contextual approximate reasoning,
- necessity of beam searching in order to keep track of several reasoning lines in parallel,
- interface between numeric and symbolic processing, i.e. determining on the speech wave the acoustic-phonetic features that constitute the fact base on which the symbolic processing operates.

Several hundreds of rules are necessary in order to cover the entire domain for a given language. Most of these rules refer to the context (i.e. the neighbourhood) of the speech segment to be identified. The following is an example of rule from the APHODEX (Acoustic PHOnetic Decoding EXpert) system developed at CRIN :

IF Right Context = /i/
AND Formant 3 is increasing
AND Formant 2 is increasing
AND First Formant visible above 1 000 Hz is at 2 200 hz
AND No discontinuity with Formant 2 of /i/
THEN Conclude : /m/.

Rule-based reasoning is an efficient means for compensating some of the errors made by more classical, algorithmic methods such as Markov models. More generally the adjunction of a certain amount of declarative processing (e.g. with production rules or logic programming) to an efficient procedural system (e.g. a stochastic model) is a good compromise for the interpretation of complex signals |16|. That makes it also possible to enhance the performances of phone segmentation algorithms (or of image segmentation as it will be shown later on).

Signal Interpretation -
A typical example of complex signal interpretation concerns the problem of ocean surveillance signal understanding. It consists in analyzing the acoustic signals emitted by different kinds of vessels in order to localize and identify these vessels. The interpretation of such signals relies mainly on the temporal evolution of spectral analysis, even though other knowledge sources are also considered (cf. section 5).

The heuristics knowledge used by the human expert can be coded into production rules, as in the HASP/SIAP system |17|. Here is an example of a rule extracted from the knowledge base of this system :

IF A source (of noise) was lost due to fade-out in the near-past
AND Similar source started up in another frequency
AND Locations of the two sources are relatively close
THEN they are the same source with confidence of .3 .

The rules are associated with a certainty factor that is used by the approximate reasoning mechanism, in a way similar to the APHODEX system. In the case of acoustic signals the problem is complicated by the fact that several distinct sources may be superimposed. The decision rule of Dempster-Shafer |18| has been used to solve this problem.

A rule-based heuristic reasoning has also been used for the interpretation of ECG signals |19| .

Image Analysis -
Algorithmic image processing methods are now often complemented by rule-based reasoning. Practical applications concern the segmentation |20|, |21|, |61| as well as the interpretation of images |22|. Such solutions are interesting for several reasons : combinatorial explosion of solutions, non-unicity of solutions, necessity of combining perceptual data with semantic information.

The system described in |20| is a good illustration of the potentiality of AI in low-level image processing, especially for segmentation purposes. Many heuristic algorithms have been proposed so far for segmenting images. AI proposes a framework that makes it possible to represent explicitly the various pieces of knowledge enbedded in these algorithms and also to evaluate the performances of the algorithms. There exist in the system three level of rules corresponding to three different levels of inference :

- on image data (operational level),
- on the control (meta level),
- on the strategy (resolution level).

A similar partition of the knowledge-base exists in blackboard architectures like HASP/SIAP or ATOME (cf. section 5).

The biomedical field corresponds to interesting applications due to the complexity of images and to the high level of expertise for interpretation : scanner images of verterbra |23|, cytology images |24|, bidimensional electrophoresis |25|.

4.2. Object Oriented Knowledge Representation and Reasoning

The interpretation of patterns necessitates the definition of "idealized" prototype patterns with characteristics features that can be found with some degree of distortion in real data. The representation by prototypes has been proposed in this case. It derives from the notion of frames |26|, initially proposed for the representation of knowledge in natural language understanding. A frame represents some kind of structured knowledge (class, sub-class, instance). It is made up of a set of elementary items, or slots. Reasoning with frames consists in filling empty slots by using other types of knowledge, e.g. production rules or procedures which are invoked when attention is given to a particular slot.

This formalism has been used in various domains of application, especially :

ECG signals |27| -
Frames describe the different classes of primitive waveforms that can be encountered in ECG displays (QRS, P, T). The reasoning strategy builds up a symbolic interpretation of the signal by starting from characteristic points ;

speech |28|-
In this case frames represent several levels of phonetic units (phonemes, phonemic classes : vowel, plosive, etc.). The definition of a frame grammar makes it possible to define a top-down decoding strategy. Therefore the frame formalism ensures the structured representation of acoustic-phonetic knowledge and also the definition of strategies for using this knowledge ;

acoustic signals |29| -
In the INTERSENSOR for ocean surveillance frames are used for the representation of signals and of noise sources (motors, etc.) ;

images -
The use of representation structures in vision |30| is well adapted to the automatic interpretation of images and scenes. Several systems have been designed on the basis of this idea. They often mix a frame based declarative representation of objects with signal processing procedures on the one hand and symbolic structure comparison techniques on the other |31|.

Object oriented languages have recently generalized these ideas, especially with the concepts of object oriented knowledge representations. That makes it possible to structure the knowledge of a particular domain while ensuring the property inheritance within a class (for instance in the description of an object in a scene). Besides, the modularity of such representations (an object contains its own description and the methods for its manipulation) makes it easier to control the interpretation since this control can be decentralized at the level of each object |32|.

The CLASSIC tool developed at INRIA |56| for data classification and/or interpretation is based on these ideas. In this system objects are represented by frames that can be organized in tree hierarchies. In order to recognize an unknown object CLASSIC looks for a prototype that is similar to this object. A rule based approximate reasoning using possibility theory makes it possible to complete the description of the object as given by sensors.

4.3. Constraints Propagation

Understanding tasks usually need contextual reasoning. An important feature of this type of reasoning consists in taking into account the constraints due to the properties of objects and in propagating them throughout the interpretation.

Such phenomena can be found in speech or image understanding. As far as images are concerned, for instance, a segmented image can be represented by a graph. In the case of stereovision a major problem is to match the two corresponding views of a scene. This matching can be carried out by algorithms that search for graph isomorphisms |57|. Such algorithms are of polynomial complexity. Therefore AI techniques of heuristic search can be used to match the graph of an image with the relational models of the universe : heuristic graph search algorithms such as A^* have proven particularly efficient |58|. A third method has also been used for the matching operation. It is an iterative method based on association graph |59| that consists in matching regions in the two images and in propagating the constraints which appeared during this operation. This important problem of image registration in stereovision illustrates the role of AI heuristic search techniques in conjunction with classical methods.

5. MODELS FOR KNOWLEDGE SOURCES COOPERATION

5.1. Multi knowledge-based systems

We have seen that the interpretation of complex patterns or signals necessitates a bunch of various knowledge and information sources. A major problem in the design of a practical system is to define an adequate model for the cooperation of these sources. Such model should also be able to allow for multi-sensory data fusion in domains like robotics.

A primary solution consists in coding all the knowledge sources within a single framework. That leads to efficient implementations in relatively simple cases. This solution was used for speech sentence understanding in the HARPY system |33| and for image understanding in the ARGOS system |34|. In these cases all available knowledge sources are coded under the form of a single stochastic network that contains all possible forms of the signals to be interpreted. The interpretation consists in finding the best path through the network thanks to classical graph searching methods (e.g. beam search). A similar method consists of representing the variability of the signals into some stochastic models, e.g. Markov models.

The idea of integrating a certain amount of knowledge by precompilation is also often used. For instance, phonological knowledge which is responsible for phonetic alterations of words in context can be represented in terms of contextual rewriting rules which apply to word entities in the lexicon |35|. This application of phonological rules to basic word-forms in the lexicon can be done either during the understanding process at each access to the lexicon or all at once in a single, preliminary phase. This phase consists of a precompilation of the set of phonological forms of the words. This latter solution is much more efficient as far as the computation time is concerned and is used in several large systems for speech understanding, such as HWIM |36|.

Similarly, the use of some kind of precompiled knowledge increases the efficiency of vision. The MIRABELLE system |37| we have developed in our group uses a compiler of descriptions for interpreting handwritten drawings. In a preliminary phase this compiler extracts information from the structural description of the class of drawings to be recognized. This information is then used during the recognition process in order to control the analysis. We have extended this idea to the interpretation of 3D scenes in the TRIDENT vision system |38|. Precompiling a knowledge base (e.g. a set of production rules) also contributes to efficiency, as for example in the PROSPECTOR system |39|.

Generally speaking, the main problem consists in making cooperate : several sub-systems with different operating modes within the same structure. From the various experiments that were carried out (especially in automatic speech recognition) two main classes of models can be derived :

- hierarchical models organized around a supervisor which controls the application of the different knowledge sources in a predetermined order ;
- more flexible models in which the order of knowledge sources invokation is not predetermined but dynamically decided in an opportunistic way according to the state of the interpretation process. Such models will now be described in greater detail.

5.2. The blackboard model

The blackboard model enables the communication between different knowledge sources through a common data structure, called blackboard that can be viewed as a generalized fact base.

A blackboard system is usually made up of three basic elements :

- **a set of knowledge sources**, each comprising two parts :
  Condition which specifies the situations where the source can be activated.
  Action which properly defines the action carried out by the knowledge sources. The "Action" part can be either a program (procedural knowledge) or a set of production rules.

The various knowledge sources enbedded in a system are independent and ignore each other. In order that a source be invoked the facts and hypotheses already available must meet the activation conditions. This information is stored in the blackboard : knowledge sources thus communicate only through the blackboard ;

- **a blackboard**, fact base of the system, usually but not always divided in hierarchical levels corresponding to the conceptual levels of the application domain. A knowledge source can accede to one or several levels in order to emit, modify or cancel an hypothesis of interpretation. The role of the blackboard is therefore twofold :

. it ensures message passing (i.e. an hypothesis) from one knowledge source to another : every source asynchronously emits hypotheses at its own level of expertise (e.g. in speech recognition : phoneme, word, phrase, sentence) and posts them on the blackboard. A hypothesis emitted by a particular knowledge source Si may activate another source Sj. The activation of the knowledge sources is thus pattern-directed ; this declarative scheme turned out to be very fruitful in AI |40|.

. it contains all the different possible interpretations of the input data in terms of combination of hypotheses.

- **a control structure** which controls the operations according to a certain strategy. The blackboard model makes it possible for the designer to implement sophisticated control strategies that are made necessary by the complexity of understanding problems. Another advantage of the model is that the control itself can be considered as one or several particular knowledge sources.

The architecture of the ATOME blackboard development system |41| that is presently developed in Nancy is given in figure 6. ATOME is a good illustration of the principles of the blackboard model. Three levels of knowledge have been defined in this system :

- specialists (lower level) which represent the knowledge specific to an application domain. The specialists are the only knowledge sources having access to the blackboard,
- tasks (meta-level) which activate specialists according to specific events that occur in the blackboard,
- the overall strategy which activates tasks according to the context of the interpretation.

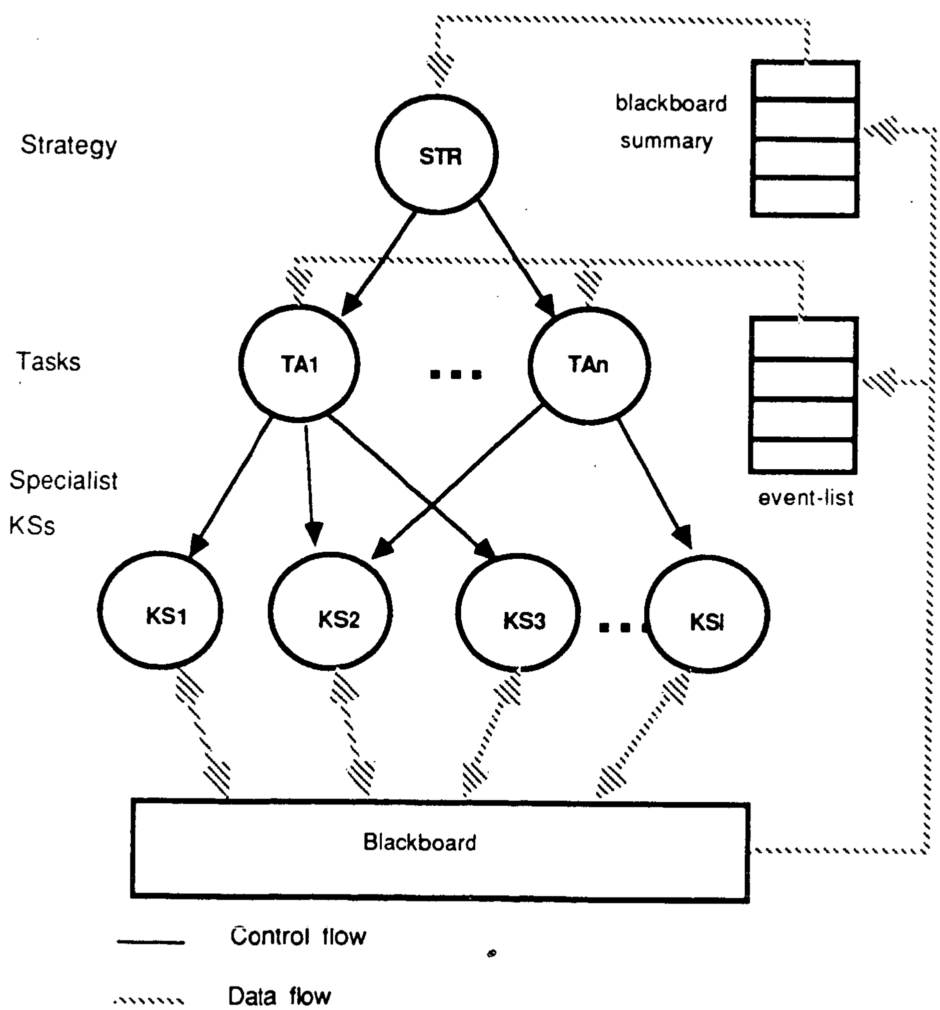

FIGURE 6
ATOME's architecture

The blackboard model was first implemented in the HEARSAY II speech understanding system |42|. In HEARSAY knowledge sources were procedural programs as well as in the VISIONS system |43| that was developed for image interpretation. The same model has then been used for underwater signal interpretation : HASP/SIAP |17|, INTERSENSOR |29| and also for the interpretation of data in cristallography : CRYSALIS |44|. The blackboard is actually well adapted to signal understanding since the knowledge and information to be taken into account in these operations are very diverse. The model proposed by Nazif and Levine in their vision system |20| is a slightly different version of the blackboard model. In their system (figure 7) various processes responsible for different understanding tasks such as system initialization, line processing, region processing, general control, etc. communicate through two different "blackboards" :

- a short term memory (STM) which contains the data and facts about the actual problem (initial data, segmentation results and final interpretation),
- a long term memory (LTM) which contains knowledge and meta-knowledge in the form of rules.

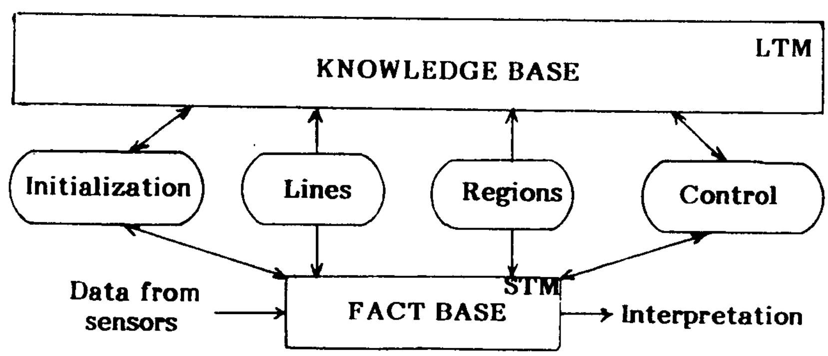

FIGURE 7
A vision system based on a two-blackboard model.

5.3. Multi-Expert Systems

The blackboard model is rather general and therefore it can sometimes be inefficient if the corresponding control structures are not sophisticated enough. Other models have been proposed for the cooperation of knowledge sources. The model of expert society has been used in various domains of signal understanding. It differs from the blackboard model at the levels of control (each expert has its own control structures) and of communication (the experts communicate directly with each other). This solution was used in speech recognition |45| and also in vision : figure 8 shows the principle of the SIGMA system |60| in which three expert systems cooperate to solve the problem of interpreting an image.

The model of specialist society |62| also features interesting properties for pattern interpretation : message communication between processors and knowledge sources, distributed control, etc.

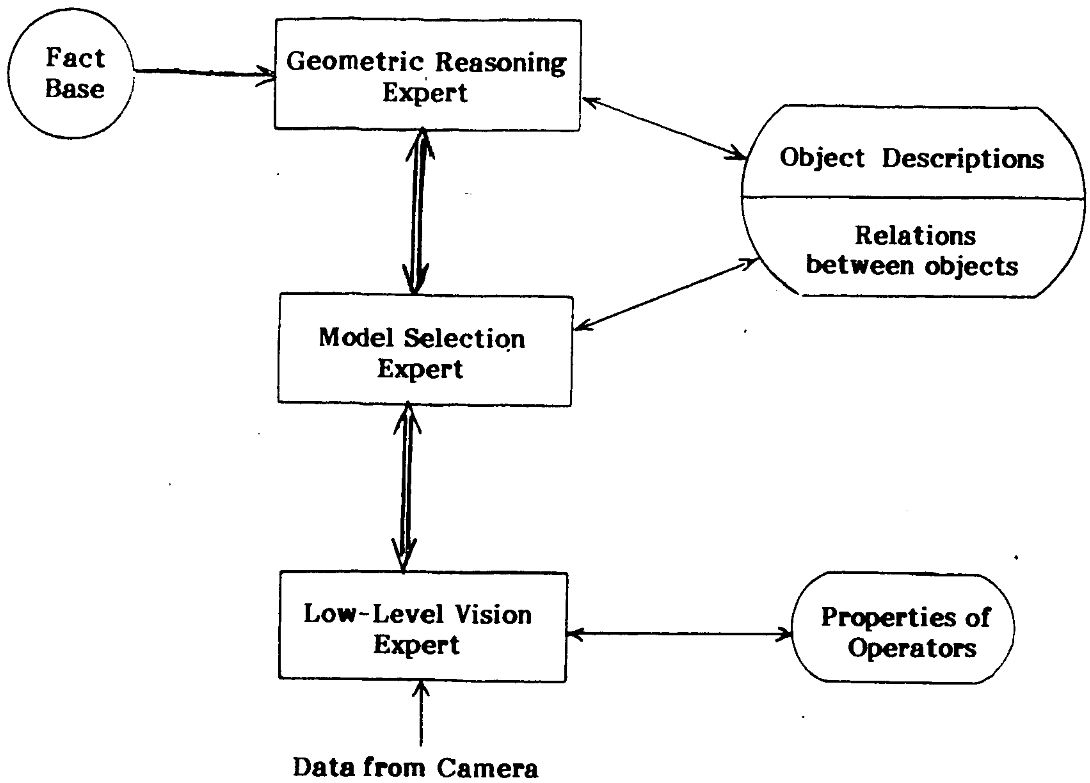

FIGURE 8
The SIGMA multi-expert vision system.

6. SEARCH STRATEGIES AND CONTROL STRUCTURES

It has been shown precedingly that in order to use various knowledge sources efficiently, it is necessary to implement sophisticated control structures for controlling the interpretation process. Moreover, the combinatorial explosion of solutions has to be reduced by means of appropriate search strategies. We will now discuss these points and show their influence on the overall architectures of understanding systems.

6.1. Models for Control Structures

One way of coping with the uncertainty inherent to understanding problems consists in using hypothesis valuation techniques and of propagating "confidence factors" during the interpretation process. The implementation of sophisticated control structures is also of primary importance. There are two basically different philosophies for control structures :

- bottom-up or data driven control, which emits hypothesis in accordance with the input data and tries progressively to build up the interpretation of the object (sentence, scene, situation, etc.) under consideration,

- a top-down or model driven control, which consists in emitting hypotheses about the input data by taking into account high level knowledge available. This type of [illegible] be associated with the classical AI notion of problem reduction.

When knowledge is coded under a Condition/Action form (like production rules or the blackboard model) these two control structure classes correspond to a forward and a backward chaining respectively.

In fact, the two methods refer to the same knowledge and manipulate the same intermediate notions (e.g. phrases, words, syllables, etc. for speech or objects, regions, lines, etc. for images). They mainly differ in the order in which the different processings are carried out and in the degree of detail into which they enter. The notions of bottom-up and top-down control are an oversimplification of the implementation of processing levels interaction in actual systems. In the human brain control structures reach still much higher level of sophistication |46|.

In a bottom-up speech recognition system, a word lattice containing all the possible occurrences of words within a sentence is built up from the acoustic data and the lexical knowledge. Syntactic and semantic knowledge are then used to select the most plausible sequence of words that will represent the actual interpretation of the sentence. This method is interesting, since it is not too sensitive to noisy data. However it becomes untractable when the size of the application (number of words, complexity of the language) increases, due to the large number of possible solutions. The same situation exists in vision : a purely bottom-up approach consisting of starting from pixels in order to build up a scene is only valuable in simple industrial applications where adequate illumination conditions make it possible to only manipulate binary images.

In top-down approaches, high level knowledge about the application universe is used for making assumptions about the constituents of a sentence or of a scene. The predictive aspect of top-down control makes it possible to cancel a large number of hypotheses. This point is particularly interesting in large applications where there is a high number of possible combinatorial solutions. On the other hand, a purely top-down control is relatively more affected by noise in the incoming data.

For simple applications, an understanding system may operate in a strictly bottom up way. On the other hand, it is very rare to find purely top down systems. An example is one version of the HWIM speech understanding system based on Klatt's analysis-by-synthesis model |47|. In this version, the top down assumptions about words are pushed down to the level of acoustic signal by means of a speech synthesis system. A word assumption is then verified by dynamic comparison of its acoustic spectrum with a position of the input sentence. This solution is not computationally efficient but it is interesting in theory. One version of the ESOPE speech understanding system |48| is also based on a similar idea.

In most cases, systems based on top-down control structures incorporate some bottomup process. The most commonly encountered arrangement consists of deriving a bottom-up symbolic description of the input data that corresponds at the same time to an important reduction of these data. This description might take the form, for example, of a phoneme lattice in the case of speech or a set of lines, regions or partially identified objects in vision. These preprocessed data are then taken into account by top-down processors.

As far as speech understanding goes, bottom-up or partially top-down systems are almost equally efficient for processing sentences of artificial, highly constrained languages (see, for instance, the MYRTILLE I system |49| for a typical example). But for pseudo-natural languages it is absolutely necessary to implement much more sophisticated control strategies that combine bottom-up and top-down methods. Such structures can be found in the systems we have already mentioned : HEARSAY II, HWIM, MYRTILLE II |50|, etc.

It is interesting to consider the development of understanding systems during the past fifteen years, both in vision and in speech. The first systems that were built relied heavily on a bottom-up methodology which consisted of extracting the maximum amount of information from the data by the use of powerful but blind techniques of signal processing and pattern recognition.

As we have already seen, except in simple cases (binary images or artificial sentences), such methods cannot handle the combinatorial explosion of possible solutions. They were therefore abandoned and replaced by top-down methods and then by mixed methods for understanding as in the case of the ACRONYM system, already mentioned. This system is based on a geometrical description of objects which is independent of the application domain. However, these models are not yet completely satisfactory and there is now a tendency towards basic studies about the physical and physiological phenomena that are involved for instance in the process of image or speech production and perception : properties of objects (luminosity, brightness, textures, geometry of surfaces, etc.), vision process |51|, etc. on the one hand, phonetic and phonological properties of sounds, phonation and audition processes, etc. on the other. This promising new tendency calls on a number of very different disciplines and necessitates the elucidation of diverse knowledge and expertise domains.

6.2. Search Strategies

At each step during the understanding process a potentially large set of partial solutions has to be examined. In order to restrict this set to a computationnally reasonable size and to avoid any combinatorial explosion, sophisticated search strategies have to be implemented so that an acceptable interpretation of the incoming data - though not always the optimal one - is finally reached. The search problem is well known in AI and it has been addressed a great number of times. Search strategies are crucial in complex problems like understanding problems or vision since the solution space is very large, whereas in more restricted tasks such as diagnostic expert systems (e.g. the MYCIN system |52| which was among the first systems developed) an exhaustive search throughout the solution space is practicable.

Dynamic programming yields an optimal solution to the search problem. For instance, it has been implemented by IBM |53| in association with a Markov model of the language by using the Viterbi algorithm. Relaxation techniques used in image processing are approximately equivalent to this algorithm.

Dynamic programming algorithms have a complexity proportional to n^2 where **n** represents the number of states to be explored. They are therefore often much too time-consuming, in which case it is necessary to use suboptimal but faster heuristic search strategies. This means that, instead of a global strategy, we have a sequence of local decisions which will hopefully lead to a global acceptable solution in accordance with the general prediction-verification paradigm. Two widely used strategies are :

- **a best-first search strategy with backtracking :** it consists in keeping the most promising solution at each step and in developing it as far as possible. If this becomes impossible, a backtracking algorithm is activated in order to consider another solution. This strategy can be found in numerous systems, since it is easy to implement and efficient for small-and medium-size applications,

- **a beam-search strategy** : this consists of simultaneously retaining the **k** best solutions and developing them in parallel, without any backtracking. This strategy is usually more efficient than the best-first strategy, especially in the case of speech understanding |54|. It is the strategy used in the HARPY system.

These two strategies are very popular in various domains of AI and have been adapted in a number of different ways. Specific constraints of speech and vision have led to the design of a new class of search strategies, called anchor-point or island-driven strategies. The basic idea consists of starting the interpretation of a sentence or an image from one or several reliable anchor-points. Such strategies are rather complex to implement but prove to be very useful for processing pseudo-natural sentences or images and handwritten sketches |55|.

7. CONCLUSION

Artificial Intelligence, and especially knowledge-based techniques, can help solving pattern and signal understanding problems typical of advanced Pattern Recognition. Several issues related to these questions have been discussed in this paper and illustrated by practical examples. The understanding of complex patterns such as speech, images, signals can be viewed as the progressive transformation of analog data extracted from the real world into a high level symbolic description. This transformation process involves numeric procedural techniques together with symbolic knowledge-based reasoning. Specific architectures have been designed for the interaction between numeric and symbolic processing and for the cooperation of various knowledge sources. Finally, this paper shows that the two fields of Artificial Intelligence and Pattern Recognition are essentially complementary for solving advanced problems of complex pattern understanding.

REFERENCES

|1| Fu,K.S., Digital Pattern Recognition (Springer Verlag, 1976).

|2| Simon, J.C., La reconnaissance des formes par algorithmes (Masson, 1985).

|3| Rabiner, L.R. et al., On the Use of Hidden Markov Models for Speaker Independent Recognition of Isolated Words from a Medium Size Vocabulary, ATT Journal, 63, (April 1984) pp. 627-642.

|4| Miclet, L., Méthodes structurelles pour la reconnaissance des formes (Eyrolles, 1984).

|5| Hummel, R.A. and S.W. Zucker, On the Foundations of Relaxation Labeling Processes, IEEE Trans. PAMI, 5, (1983) pp. 267-287.

|6| Fu, K.S., Syntactic Pattern Recognition and Applications (Prentice Hall, 1982).

|7| Haton, J.P., Intelligence artificielle en compréhension automatique de la parole : état des recherches et comparaison avec la vision par ordinateur, T.S.I., vol. 4, n° 3, (1985) pp. 265-287.

|8| Oshima, M., Shirai, Y., Object Recognition using Three-Dimensional Information, IEEE. Trans. PAMI, 5, n° 4 (1983).

|9| Ballard, D.H. and Brown, C.M., Computer Vision, (Prentice Hall, 1982).

|10| Bunke, H., Sagerer, G. and Niemann, H., Model Based Analysis of Scientific Image Sequences of the Human Heart, in : T.S. Huang, (ed.), Image Sequence Processing and Dynamic Scene Analysis (Springer Verlag, 1983).

|11| Brooks, R.A., Symbolic Reasoning Among 3D Objects and 2D Models, Artificial Intelligence, 17 (1-3) (1981) pp. 285-348.

|12| Zue, V. and Cole, R., Experiments in Spectrogram Reading, Proc. IEEE ICASSP-79, Washington (1979) pp. 116-119.
|13| Carbonell, N. Fohr, D., Haton, J.P., Lonchamp, F., Pierrel, J.M., An Expert System for the Automatic Reading of French Spectrograms, Proc. IEEE ICASSP-84, San Diego (1984) pp. 48.8.1.-48.8.4.
|14| Gillet, D. et al., SERAC : un système expert en reconnaissance acoustico-phonétique, 4ème congrès RF-IA (Jan. 1984).
|15| Memmi, D., Eskenazi, M., Mariani, J., Un système expert pour la lecture de sonagrammes, Speech Communication, 2 (2-3) (1983) pp. 234-236.
|16| Haton, J.P., Carbonell, N., Fohr, D., Mari, J.F., Kriouile, A., Interaction between Stochastic Modeling and Knowledge-Based Techniques in Acoustic-Phonetic of Speech", Proc. IEEE ICASSP-87, Dallas, Texas (April 1987).
|17| Nii, H.P. et al., Signal-to-Symbol Transformation : HASP/SIAP Case Study, AI Magazine, (1982) pp. 23-25.
|18| Hughes, R.C., Maksym, J.M., Acoustic Signal Interpretation : Reasoning with Non-Specified and Uncertain Information, Pattern Recognition, 18, n° 6 (1985) pp. 475-483.
|19| Birman, K.P., Rule-Based Learning for more accurate ECG Analysis, IEEE Trans. PAMI, 4 (1981) pp. 369-380.
|20| Nazif, A.M. and Levine, M.D., Low Level Image Segmentation : An Expert System, IEEE Trans. PAMI, 6 (5) (1984) pp. 555-577.
|21| Zhang, Z., Simaan, M., A Rule-Based Interpretation System for Segmentation of Seismic Images, Pattern Recognition, 20, n° 1 (1987) pp. 45-53.
|22| Mc Keown, D.M., Jr. and Mc Dermott, J., Toward Expert Systems for Photo Interpretation, IEEE Conf. on Trends and Applications-83, (1983) pp. 33-39.
|23| Vesoul, P. et al., ORFEVRE (Outil de Reconnaissance des Formes et d'Expertise sur les Vertèbres pour la RadiologiE) : système expert et image, proposé au 6ème congrès RF-IA, Antibes (1987).
|24| Garbay, C., Segmentation d'images : connaissance et stratégies d'analyse, 5ème congrès RF-IA, Grenoble (1985).
|25| Appel, R. et al., A Computerized System for Spot Detection and Analysis of Two-Dimensional Electrophoresis Images, in : Thalmann (ed.), Computer-Generated Images (Springer Verlag, 1985).
|26| Schank, R.C., Conceptual information Processing (North Holland, 1975).
|27| Lee, H.S. and Thakor, N.V., Frame-Based Understanding of ECG Signals, Proc. First Conf. on Applic. of Artificial Intelligence, Denver (1984).
|28| Haton, J.P. Damestoy, J.P., A Frame Language for the Control of Phonetic Decoding in Continuous Speech Recognition, Proc. IEEE ICASSP-85, Tampa, Florida (1985) pp. 1565-1568.
|29| Maksym, J.N. et al., Machine Analysis of Acoustical Signals, Pattern Recognition, 16, n° 6 (1983) pp. 615-625.
|30| Marr, D., Vision (Freeman, 1982).
|31| Kak, A.C. et al., Knowledge-Based Robotics, IEEE Conf. on Robotics and Automation (1987).
|32| Kim, J.H., Payton, D.W. and Olin, K.E., An Expert System for Object Recognition in Natural Scenes, Proc. First Conf. on Apllic. of Artificial Intelligence, Denver (1984).
|33| Lowerre, B.T., The HARPY Speech Recognition System, Techn. Report, Carnegie-Mellon University, Department of Computer Science (1976).
|34| Rubin, S., The ARGOS Image Understanding System, Ph. D. Thesis, Carnegie-Mellon University, Pittsburgh, Pa. (1978).
|35| Shoup, J.E., Phonological Aspects of Speech Recognition, in : Lea, W., (ed.), (Prentice Hall, 1980).
|36| Woods, W.A. et al., Speech Understanding Systems, Final Technical Progress Report, Report n° 3438, Vol. I-V, BBN (1976).

|37| Masini, G., Réalisation d'un système de reconnaissance structurelle et d'interprétation de dessins, Thèse de 3ème cycle, Université de Nancy I (1978).

|38| Masini, G. et F. Zaroli, Présentation de TRIDENT : un système d'interprétation d'images tridimensionnelles, 4ème congrès RF-IA, AFCET-INRIA, Paris (1984).

|39| Konolige, K., An Inference Net Compiler for the PROSPECTOR Rule-Based Consultation System, Proc. IJCAI-6 (1979).

|40| Waterman, J. and F. Hayes-Roth, (eds.), Pattern-Directed Inference Systems (Academic Press, 1978).

|41| Lâasri, H., B. Maître, J.P. Haton, ATOME, outil d'aide au développement de systèmes multi-experts, submitted at 6th Cong. Pattern Recognition and Artificial Intelligence, Antibes (1987).

|42| Lesser, V.R. et al., Organization of the HEARSAY II Speech Understanding System, IEEE Trans. ASSP, 23 (1) (1975) pp. 11-23.

|43| Hanson, A.R. and E.M. Riseman, VISIONS : A Computer System for Interpreting Scenes, in : A.R. Hanson and E.M. Riseman, (eds.), Computer Vision Systems, (Academic Press, 1978).

|44| Engelmore, R.E. and A. Terry, Structure and Function of the CRYSALIS System, Proc. IJCAI-6 (1979).

|45| De Mori, R., P. Laface, G. Petrone and M. Segnan, Access to a Large Lexicon Using Phonetic Features, EUSIPCO (Erlangen, 1983).

|46| Geschwind, N., Neurological Knowledge and Complex Behaviors, Cognitive Science, 4, n° 2 (1980) pp. 185-193.

|47| Klatt, D.H., Review of the ARPA Speech Understanding System, in : Reddy, D. (ed.), Speech recognition (1975).

|48| Mariani, J., ESOPE : un système de compréhension de la parole continue, Thèse d'Etat, Université de Paris VI (1982).

|49| Pierrel, J.M., Contribution à la compréhension automatique du discours continu, Thèse de 3ème cycle, Université de Nancy I (1975).

|50| Haton, J.P. and J.M. Pierrel, Data-Structures and Architecture of MYRTILLE II Speech Understanding System, 4th Int. Joint Conf. on Pattern Recognition, Kyoto (1978).

|51| Brady, J.M., The Changing Shape of Computer Vision, Artificial Intelligence, 17 (1981) pp. 1-3.

|52| Shortliffe, E.H., Computer-Based Medical Consultation : MYCIN (American Elsevier, 1976).

|53| Jelinek, F., Self-Organized Continuous Speech Recognition, in : Haton, J.P., (ed.), Automatic Speech Analysis and Recognition (D. Reidel, 1982).

|54| Quinton, P., Utilisation de contraintes syntaxiques pour la reconnaissance de la parole continue, TSI, 1, n° 3 (1982) pp. 233-248.

|55| Haton, J.P. and R. Mohr, A New Parsing Algorithm for Imperfect Patterns and its Applications, Proc. ICPR, San Diego (1976).

|56| Granger, C., Fuzzy Reasoning in a Knowledge-Based System for Object Classification, Proc. ECAI, Brighton (1986) pp. 163-170.

|57| Ulmann, J.R., An Algorithm for Subgraph Isomorphism, JACM, 23, n° 1 (1976) pp. 31-42.

|58| Bunke, H. and Allerman, G., Inexact Graph Matching for Structural Pattern Recognition, Pattern Recognition Letters, 1 (1983) pp. 245-253.

|59| Ambler, A.P. et al., A Versatile Computer Controlled Assembly System, Artificial Intelligence, 6, n° 2 (1975) pp. 129-156.

|60| Matsuyama, T. and V. Hwang, SIGMA : A Framework for Image Understanding, Proc. IJCAI-9 (1979).

|61| Stansfield, S., ANGY : A Rule-Based Expert System for Automatic Segmentation of Coronary Vessels from Digital Substracted Angiograms, IEEE trans. PAMI, 5, (2) (1986) pp. 188-189.

|62| Gong, Y. and Haton, J.P., Une société de spécialistes à niveaux multiples pour l'interprétation de signaux, proposé au congrès AFCET, RF-IA, Antibes, (1987).